SUPERVISED LEARNING CALIBRATION OF AN ATMOSPHERIC LIDAR

Adi Vainiger[1], *Omer Shubi*[2], *Yoav Y. Schechner*[1]
Yin Zhenping[3], *Holger Baars*[4], *Birgit Heese*[4], *Dietrich Althausen*[4]

[1]Viterbi Faculty of Electrical and Computer Engineering,
Technion - Israel Institute of Technology, Haifa, Israel
[2]Faculty of Industrial Engineering and Management,
Technion - Israel Institute of Technology, Haifa, Israel
[3]School of Remote Sensing and Information Engineering Wuhan University, China
[4]Leibniz Institute for Tropospheric Research (TROPOS), Leipzig, Germany

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883078

ABSTRACT

Calibration of an atmospheric lidar is often required due to variations in the electro-optical system. Rayleigh fitting commonly performed may fail under various conditions. Temporal and spatial variations both affect lidar signals. We hence opt for spatiotemporal analysis. We present a novel deep-learning (DL) lidar calibration model based on convolutional neural networks (CNN). We demonstrate our method on simulated data that mimics natural ground-based pulsed time-of-flight lidar signals. Such an approach can better address measurements with a poor signal-to-noise ratio (SNR) and provide a more frequent calibration.

Index Terms— Lidar calibration, CNN, DL

1. INTRODUCTION

A lidar [1] is affected by external factors and internal wear, such as in-door temperature fluctuations affecting optical alignment and laser energy. Consequently, lidar calibration is often required due to variations in the electro-optical system. Calibration can be performed under stable conditions, excluding times when the lidar's cabin temperature fluctuates rapidly, when fog or clouds appear, or during maintenance [2, 3]. This hinders lidar analysis.

A standard calibration method for valid periods employs Rayleigh fitting [4]. This method searches for a reference height–range in which aerosol–free conditions occur. At these altitudes, aerosol scattering is negligible in comparison with Rayleigh scattering. Rayleigh fitting includes spatial and temporal averaging operations, leading to information loss and possible inaccurate estimations of the lidar constant (LC). Additionally, this method may break during complex atmospheric conditions, e.g., when several aerosol layers are present [5]. Typically, this leads to searching higher altitudes for a reference height–range. However, the SNR drops with altitude. Poor SNR conditions, weak aerosol signals, and imperfect cloud detection can lead to incorrect reference heights and, thus, inaccurate LC estimations.

Over a day, a lidar continuously samples vertical profiles in the atmosphere, creating a two-dimensional (2D) spatiotemporal map. Hence, we hypothesize that analysis should address dynamic phenomena of both the atmosphere and the lidar system. This can be achieved by employing spatiotemporal approaches. Here we demonstrate this concept to benefit lidar calibration. Recent works [6, 7] suggest employing sequential methods applied to consecutive profile measurements. We advocate adopting powerful and highly developed learning-based image processing. Such methods may better overcome challenges posed by sparse or low-SNR signals [8–10].

We develop a neural network (NN) model for an atmospheric lidar calibration while employing CNN layers. A CNN-based model applied to 2D lidar data can capture rich spatiotemporal features, improve SNR, and reduce information loss. Our goal is to avoid a prior reference height search for lidar calibration. A second objective is developing a robust model by considering long- and short-term variations of lidar measurements during training. A model meeting these goals may provide a more frequent and accurate calibration.

Achieving these goals requires a large amount of annotated data. Yet, databases with frequent lidar calibration are rare. Inspired by the recent contribution of synthesized databases to DL methods in various disciplines, we suggest augmenting data for our objective. Expert knowledge and statistical information gained over the years enable us to generate traceable data while mimicking natural lidar signals. We achieved this using the Atmospheric Lidar Data

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810370: CloudCT). Yoav Schechner is the Mark and Diane Seiden Chair in Science at the Technion. He is a Landau Fellow - supported by the Taub Foundation. His work was conducted in the Ollendorff Minerva Center. Minvera is funded through the BMBF.
(Email:addalin@campus.technion.ac.il)

Augmentation (ALiDAn) framework [11], which follows the physical model of ground-based pulsed time-of-flight lidars. The generated data correspond to the elastic channels of the Polly^{XT} lidar [12] by TROPOS; further details in Section 3.1. This work explores several NN configurations to achieve the above joint goals. We examine the contribution of different data sources that feed the model. We compare the model for different wavelengths and seasons while testing various hyper-parameters. Finally, we conclude with current results on simulated data and discuss potential future directions.

2. LIDAR CALIBRATION - LEARNING MODEL

Let samples of a spatiotemporal lidar signal be arranged in an array with dimensions of $m \times n$, representing height and time bins, respectively. Spatiotemporal particle attenuated backscatter coefficient due to elastic interactions with the lidar laser beam is $\boldsymbol{\beta}_{\text{ATTN}} \in \mathbb{R}_{+}^{m \times n}$. Set diagonal matrices to multiply $\boldsymbol{\beta}_{\text{ATTN}}$: the height-dependent square values of altitude bins and the overlap function set respectively the diagonals of $\mathbf{R}^2 \in \mathbb{R}_{+}^{m \times m}$ and $\mathbf{O} \in \mathbb{R}_{+}^{m \times m}$, and the time-dependent LC values set the diagonal of $\mathbf{P}^{\text{LC}} \in \mathbb{R}_{+}^{n \times n}$. Let $\mathbf{P}^{\text{BG}}$ represent the background (BG) signal from sunlight scatter. Then the array-based formalism of a spatiotemporal lidar signal, at an arbitrary wavelength λ, following the model in [1], is

$$\mathbf{P} = \mathbf{R}^{-2}\mathbf{O}\boldsymbol{\beta}_{\text{ATTN}}\mathbf{P}^{\text{LC}} + \mathbf{P}^{\text{BG}} \in \mathbb{R}_{+}^{m \times n}\,[\text{photons}]. \quad (1)$$

Let describe the supervised learning model of lidar calibration. The model is fed by input $\mathbf{X}$ of size $h_{\text{IN}} \times w_{\text{IN}} \times c_{\text{IN}}$, set by c_{IN} concatenated 2D maps each of size $h_{\text{IN}} \times w_{\text{IN}}$. Here h_{IN}, w_{IN}, and c_{IN} are respectively the altitude, time, and channel dimensions of $\mathbf{X}$. We suggest that each sample $\mathbf{X}$ be composed of a 2D lidar measurement and information derived from additional sources to enable capturing natural statistics. The suggested sources to form the input $\mathbf{X}$ are presented in Fig. 1 and described next.

The first channel is the spatiotemporal range–corrected lidar measurement $\mathbf{X}_1 = \mathbf{P}_{\text{RC}} = \mathbf{R}^2\mathbf{P}[\text{photons} \cdot \text{km}^2]$. To avoid possible biases, we do not subtract $\mathbf{P}^{\text{BG}}$ from lidar measurements ; unlike a common lidar pre-processing phase. The second channel represents the molecular attenuated backscatter coefficient $\mathbf{X}_2 = \boldsymbol{\beta}_{\text{ATTN}}^{\text{mol}}[\frac{1}{\text{km} \cdot \text{sr}}]$. Such input can be obtained from meteorological measurements. The attenuated backscattered signal is commonly divided by an aerosol transmittance factor [2]; however, its prior–assessment may not always be available. Hence, we do not use the aerosol factor.

We examine two models, which differ by c_{IN}. The first model uses $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$. The second model adds a third channel that represents background signal information. For the latter, we suggest two different modes; using the averaged BG photon counts $\mathbf{X}_3^{\text{v1}} = \mathbf{P}^{\text{BG}}[\text{photons}]$, or the range corrected BG signal $\mathbf{X}_3^{\text{v2}} = \mathbf{R}^2\mathbf{P}^{\text{BG}}[\text{photons} \cdot \text{km}^2]$.

A lidar measurement $\mathbf{P}$ has a very large dynamic range. To highlight information of interest, we suggest using a transformation similar to gamma correction, common in image processing. Let $\boldsymbol{\gamma}_x$ be a vector with a scalar value per channel. We define a point-wise power transform, at a rate set by $\boldsymbol{\gamma}_x$ to each input channel before feeding to the NN

$$\tilde{\mathbf{X}} = \mathbf{X}^{\boldsymbol{\gamma}_x}. \quad (2)$$

Let $\mathbf{F}$ be a model with a set of learned weights $\boldsymbol{\Theta}$, to estimate the LC denoted as $\hat{\mathbf{Y}}$, for a given input $\mathbf{X}$, then

$$\hat{\mathbf{Y}} = \mathbf{F}(\tilde{\mathbf{X}}, \boldsymbol{\Theta}). \quad (3)$$

The weights $\boldsymbol{\Theta}$ are calculated during a training process by minimising the LC errors of model predictions $\mathbf{F}(\mathbf{X}, \boldsymbol{\Theta})$ relative to $\mathbf{Y}$ in a gradient decent manner

$$\min_{\boldsymbol{\Theta}} \text{Loss}(\mathbf{Y}, \hat{\mathbf{Y}}). \quad (4)$$

Where $\mathbf{Y}$ is a vector containing the corresponding ground–truth LC values of samples in $\mathbf{X}$.

The model contains four CNN layers and two Fully Connected (FC) layers, presented in Fig. 2(a). Each CNN hidden layer calculates n_{H} feature maps. The feature maps dimensions' are reduced in the time domain by 2 and in the height domain by 4 until reaching a feature map of size $h_{\text{OUT}} \times w_{\text{OUT}}$. Each CNN layer includes (I) 2D convolution kernels that calculate a feature map of the layer. (II) Batch Norm operation for standardizing the learned features. (III) A Rectified Linear Unit (ReLU) activation function. (IV) Dropout that randomly "shuts down" $10\% - 15\%$ of the neurons during training to avoid overfitting. (V) MaxPool, a pooling layer, taking maximal values and reducing feature dimensionality.

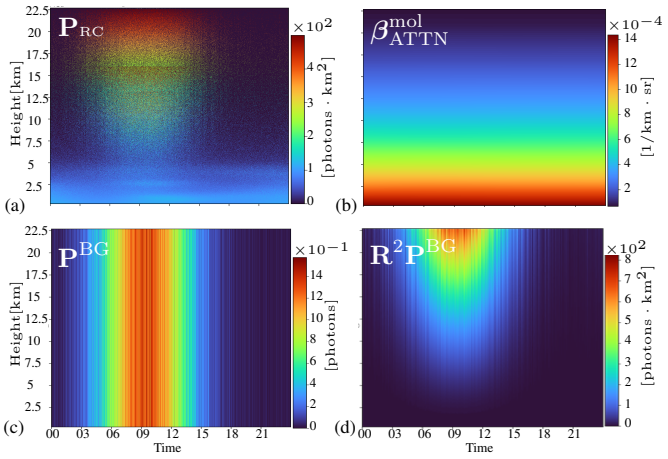
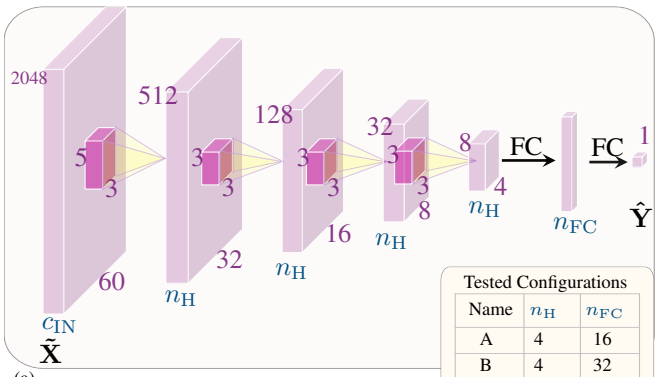


Fig. 1. Suggested input sources to the model: (a) $\mathbf{X}_1 = \mathbf{P}_{\text{RC}}$, (b) $\mathbf{X}_2 = \boldsymbol{\beta}_{\text{ATTN}}^{\text{mol}}$, (c) $\mathbf{X}_3^{\text{v1}} = \mathbf{P}^{\text{BG}}$, and (d) $\mathbf{X}_3^{\text{v2}} = \mathbf{R}^2\mathbf{P}^{\text{BG}}$

3. EXPERIMENTS & RESULTS

3.1. Simulated Databases

Using the ALiDAn framework [11], we simulate natural lidar measurements over four months for two different sea-



Tested Configurations

Name	n_H	n_{FC}
A	4	16
B	4	32
C	6	16
D	6	32

Fig. 2. (a) The lidar learning calibration model, for setting $h_{IN} \times w_{IN}$ to be 2048×60. (b) The tested configurations.

sons. Statistics for generation rely on data from the literature and measurements collected in a field campaign [13] during 2016–2019; in which, a Polly^{XT} lidar and an AERONET sunphotometer were mounted on the rooftop of the Meyer Building (EE Faculty), Technion, Haifa, Israel. The simulated lidar signals are derived from a Poissonian distribution to induce natural statistics of photo–electron signals. ALiDAn allows considering different overlap functions. Here we set it to 1; we plan including overlap function in future work.

The initial database $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ represents diurnal lidar simulations for September–October 2017, simulating 61 days, with a similar amount of data per $\lambda \in [355, 532, 1064][\mathrm{nm}]$. The extended database $\boldsymbol{\mathcal{D}}_{\mathrm{E}}$ consists of $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ and a similar simulated period for April–May 2017. The latter database is challenging since it contains two seasons and a variety of system values. Data statistics are presented in Fig. 3(a).

3.2. Architecture settings

Each input size, defined as $h_{IN} \times w_{IN}$, is 2048×60 pixels, representing a 30–minute measurement of the atmosphere up to an altitude of $\sim 15.6[\mathrm{km}]$. These settings correspond to the analysis done by the PollyNet Processing Chain [3] by TROPOS. Thus, each simulated month in each database results in 4392 samples. According to the NN definitions previously set, the last feature map is of size $h_{OUT} \times w_{OUT} = 8 \times 4$.

We test two models, which differ by the number of channels c_{IN} of the input layer, earlier presented in Section 2. We conduct several experiments to determine which architecture of the suggested models achieves the best results. Additionally, for each model we test four different NN architectures $\{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$. The architectures are presented in Fig. 2(b).

3.3. Training and Evaluation

Let s be a sample index in an input batch of size n_{B}. The training loss in each iteration is

$$\mathrm{Loss} = \frac{1}{n_{\mathrm{B}}} \sum_{s=1}^{n_{\mathrm{B}}} \left| \mathrm{Y}_s - \hat{\mathrm{Y}}_s \right|. \tag{5}$$

LC values, which set $\mathbf{Y}$, vary by time and wavelength, as shown in Fig. 3(a); hence affecting the loss values. This may result in a wide range of losses, making it challenging to compare different wavelengths. Therefore, we use the Mean Absolute Relative Error (MARE) on the validation set, to fairly compare between the experiments

$$\mathrm{MARELoss} = \frac{1}{n_{\mathrm{B}}} \sum_{s=1}^{n_{\mathrm{B}}} \frac{\left| \mathrm{Y}_s - \hat{\mathrm{Y}}_s \right|}{\mathrm{Y}_s}. \tag{6}$$

Next we present the main highlights of an in-depth analysis that tests the effectiveness of the network by different aspects.The results correspond for the validation set, which is 16% randomly selected from $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ and $\boldsymbol{\mathcal{D}}_{\mathrm{E}}$. Additional experiments settings are: $n_{\mathrm{B}} = 32$ and learning rate of 2×10^{-3}.

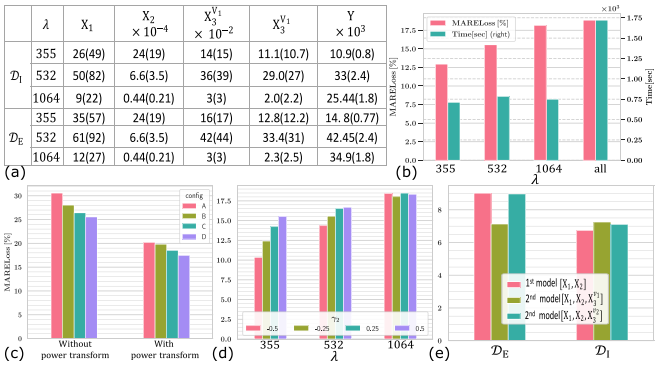
3.4. Results Analysis

Fig. 3(c) presents the average loss of the 1^{st} model, testing various power transforms, such as $\gamma_1 = 0.5$ and $\gamma_2 = \pm 0.25, \pm 0.5$, for $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ at all wavelengths. Power transform yields better learning, reducing errors from $30\% - 40\%$ to less than 10%. Testing all suggested models and configurations on $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ and $\boldsymbol{\mathcal{D}}_{\mathrm{E}}$ achieved similar results.

Data statistics vary greatly between input channels and between wavelengths. E.g., $\mathbf{X}_1$ values at 1064 are several times lower than $\mathbf{X}_1$ at 355 and 532, and are several orders of magnitude larger than $\mathbf{X}_2$ at 1064, as shown in the statistics table in Fig. 3(a). Fig. 3(b) shows that separating the model per wavelength improves accuracy and reduces the training time. Moreover, model predictions improve by using different γ values for each channel per wavelength. This is evident and consistent in many experiments and is illustrated in Fig. 3(d).

Fig. 3(e) presents comparisons for $\lambda = 355[\mathrm{nm}]$, testing the performance on long-term variations, i.e., $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ vs. $\boldsymbol{\mathcal{D}}_{\mathrm{E}}$; and BG data contribution, comparing the 1^{st} vs. the 2^{nd} models. For the 1^{st} model, training the NN on $\boldsymbol{\mathcal{D}}_{\mathrm{I}}$ gained better results. However, the contribution of the BG information, i.e., the 2^{nd} model, was more significant for training on $\boldsymbol{\mathcal{D}}_{\mathrm{E}}$; when using $\mathbf{X}_2^{\mathrm{v1}}$ the second model reached a minimum error of 7%. Similar comparisons were made for the rest of the wavelengths.

The NN performs best for $\lambda = 355[\mathrm{nm}]$. It can be expected since typically, at $355[\mathrm{nm}]$, the SNR is higher, and the backscatter of air molecules is more significant. For $532[\mathrm{nm}]$, the NN reaches 10% error; here, the backscatter of aerosols and air molecules have a similar order of magnitude. For $1064[\mathrm{nm}]$, the NN reached a minimum error of 15%, which



	λ	X_1	X_2 $\times 10^{-4}$	$X_3^{V_1}$ $\times 10^{-2}$	$X_3^{V_1}$	Y $\times 10^{3}$
$\mathcal{D}_\mathrm{I}$	355	26(49)	24(19)	14(15)	11.1(10.7)	10.9(0.8)
	532	50(82)	6.6(3.5)	36(39)	29.0(27)	33(2.4)
	1064	9(22)	0.44(0.21)	3(3)	2.0(2.2)	25.44(1.8)
$\mathcal{D}_\mathrm{E}$	355	35(57)	24(19)	16(17)	12.8(12.2)	14. 8(0.77)
	532	61(92)	6.6(3.5)	42(44)	33.4(31)	42.45(2.4)
	1064	12(27)	0.44(0.21)	3(3)	2.3(2.5)	34.9(1.8)

Fig. 3. (a) Statistics of the generated databases - $\boldsymbol{\mathcal{D}}_\mathrm{I}$, and $\boldsymbol{\mathcal{D}}_\mathrm{E}$. (b) Wavelength separation comparison for $\boldsymbol{\mathcal{D}}_\mathrm{I}$: left - average $\mathrm{MARELoss}$, right - average train time. (c) Comparison of NN configuration and power transform for $\boldsymbol{\mathcal{D}}_\mathrm{I}$ at all wavelengths. (d) Average $\mathrm{MARELoss}$ of the 1^{st} model for $\gamma_1 = 0.5$ and varying γ_2 values, tested on $\boldsymbol{\mathcal{D}}_\mathrm{I}$. (e) Comparing the background contribution at $355[\mathrm{nm}]$ and configuration D, between the generated databases - $\boldsymbol{\mathcal{D}}_\mathrm{I}$, and $\boldsymbol{\mathcal{D}}_\mathrm{E}$.

may be explained by typical low SNR conditions; due to a higher sensor sensitivity and atmospheric absorption of water particles, weakening the backscatter signal at $1064[\mathrm{nm}]$.

4. DISCUSSION & CONCLUSIONS

Overall, our model achieves $7\% - 15\%$ $\mathrm{MARELoss}$ on the simulated validation set, showing great potential to improve current lidar calibrations [2, 3]. We now consider a transfer learning phase with additional statistical variations, including various overlap functions and adapting the suggested approach to raw measurements.

5. REFERENCES

[1] Claus Weitkamp, *Lidar: range-resolved optical remote sensing of the atmosphere*, vol. 102, Springer Science & Business, 2006.

[2] Holger Baars, Patric Seifert, Ronny Engelmann, and Ulla Wandinger, "Target categorization of aerosol and clouds by continuous multiwavelength-polarization lidar measurements," *Atmospheric Measurement Techniques*, vol. 10, no. 9, pp. 3175–3201, 2017.

[3] Zhenping Yin, Holger Baars, Patric Seifert, and Ronny Engelmann, "Automatic lidar calibration and processing program for multiwavelength Raman polarization lidar," in *EPJ Web of Conferences*. EDP Sciences, 2020, vol. 237.

[4] Volker Freudenthaler, "Lidar Rayleigh–fit criteria," in *EARLINET-ASOS 7th Workshop*, 2009.

[5] Dietrich Althausen, Silke Mewes, Birgit Heese, Julian Hofer, Yoav Schechner, Amit Aides, and Vadim Holodovsky, "Vertical profiles of dust and other aerosol types above a coastal site," in *E3S Web of Conferences*. EDP Sciences, 2019, vol. 99, p. 02005.

[6] Nishant Kumar, Kirti Soni, and Ravinder Agarwal, "Prediction of temporal atmospheric boundary layer height using long short-term memory network," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 73, no. 1, pp. 1–14, 2021.

[7] Dorsa Ziaei, Jennifer Sleeman, Milton Halem, Vanessa Caicedo, Ruben Mann Delgado, and Belay Demoz, "Convolutional lstm for planetary boundary layer height (pblh) prediction," *UMBC Joint Center for Earth Systems Technology (JCET)*, 2021.

[8] Erol Cromwell and Donna Flynn, "Lidar cloud detection with fully convolutional networks," in *Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 619–627.

[9] Willem J Marais, Robert E Holz, Yu Hen Hu, Ralph E Kuehn, Edwin E Eloranta, and Rebecca M Willett, "Approach to simultaneously denoise and invert backscatter and extinction from photon-limited atmospheric lidar observations," *Applied Optics*, vol. 55, no. 29, pp. 8316–8334, 2016.

[10] Zhuofan Zheng, Weibiao Chen, Yupeng Zhang, Sijie Chen, and Dong Liu, "Denoising the space-borne high-spectral-resolution lidar signal with block-matching and 3D filtering," *Applied Optics*, vol. 59, no. 9, pp. 2820–2828, 2020.

[11] Adi Vainiger, Omer shubi, Yoav Y. Schechner, Yin Zhenping, Holger Baars, Birgit Heese, and Dietrich Althausen, "ALiDAn: Spatiotemporal and multi–wavelength atmospheric lidar data augmentation," *In review*, 2022.

[12] Ronny Engelmann, Thomas Kanitz, Holger Baars, Birgit Heese, Dietrich Althausen, Annett Skupin, Ulla Wandinger, Mika Komppula, Iwona S. Stachlewska, Vassilis Amiridis, Eleni Marinou, Ina Mattis, Holger Linné, and Albert Ansmann, "The automated multiwavelength Raman polarization and water-vapor lidar PollyXT: The neXT generation," *Atmospheric Measurement Techniques*, vol. 9, no. 4, pp. 1767–1784, 2016.

[13] Amit Aides, Aviad Levis, Vadim Holodovsky, Yoav Y Schechner, Dietrich Althausen, and Adi Vainiger, "Distributed sky imaging radiometry and tomography," in *International Conference on Computational Photography (ICCP)*. IEEE, 2020, pp. 1–12.